



# Estimating parameters of a multiple autoregressive model by the modified maximum likelihood method

Özlem Türker Bayrak<sup>a,\*</sup>, Ayşen D. Akkaya<sup>b</sup>

<sup>a</sup> Department of Industrial Engineering, Çankaya University, Ankara, Turkey

<sup>b</sup> Department of Statistics, Middle East Technical University, Ankara, Turkey

## ARTICLE INFO

### Article history:

Received 9 April 2008

Received in revised form 9 September 2009

### Keywords:

Autoregression

Student's *t*

Generalized Logistic

Modified likelihood

Non-normality

## ABSTRACT

We consider a multiple autoregressive model with non-normal error distributions, the latter being more prevalent in practice than the usually assumed normal distribution. Since the maximum likelihood equations have convergence problems (Puthenpura and Sinha, 1986) [11], we work out modified maximum likelihood equations by expressing the maximum likelihood equations in terms of ordered residuals and linearizing intractable nonlinear functions (Tiku and Suresh, 1992) [8]. The solutions, called modified maximum estimators, are explicit functions of sample observations and therefore easy to compute. They are under some very general regularity conditions asymptotically unbiased and efficient (Vaughan and Tiku, 2000) [4]. We show that for small sample sizes, they have negligible bias and are considerably more efficient than the traditional least squares estimators. We show that our estimators are robust to plausible deviations from an assumed distribution and are therefore enormously advantageous as compared to the least squares estimators. We give a real life example.

© 2009 Elsevier B.V. All rights reserved.

## 1. Introduction

A multiple autoregressive model is

$$y_t = \mu + \phi y_{t-1} + \sum_{j=1}^q \delta_j (x_{j,t} - \phi x_{j,t-1}) + e_t \quad (1 \leq t \leq n), \quad (1.1)$$

where

$y_t$  = observed value of a random variable  $y$  at time  $t$ ,

$x_{j,t}$  = value of the  $j$ th nonstochastic design variable at time  $t$ ,

$\phi$  = autoregressive coefficient,  $-1 < \phi < 1$ .

Estimation of parameters in (1.1) is perceived to be a difficult problem because estimation of  $\mu$ ,  $\delta_j$  ( $j = 1, 2, \dots, q$ ) and  $\sigma$  involves the parameter  $\phi$  and estimation of the latter is problematic [1,2]. Traditionally,  $e_t$  have been assumed to be normal  $N(0, \sigma^2)$ . There is now a realization that non-normal distributions are more prevalent in practice. It is, therefore, very important to provide solutions to the estimation problems under non-normality of the error distribution. We consider three different types of non-normal distributions which represent a very large variety of non-normal distributions and

\* Corresponding address: Çankaya Üniversitesi, Endüstri Mühendisliği Bölümü, Öğretmenler Caddesi, No:14, 06530, 100.Yıl, Ankara, Turkey.  
E-mail address: [ozlemt@cankaya.edu.tr](mailto:ozlemt@cankaya.edu.tr) (Ö.T. Bayrak).

are amenable to modified maximum likelihood estimation [3]: (i) long-tailed symmetric (LTS), (ii) skew distributions represented by the Generalized Logistic, and (iii) short-tailed symmetric (STS) distributions; see also [3]. Since the maximum likelihood estimators (MLEs) are elusive for these families, we derive the modified maximum likelihood estimators (MMLEs). Under some very general regularity conditions, the MMLEs are known to be asymptotically fully efficient (i.e., they are unbiased and have minimum variances); see [4]. For small  $n$ , they are known to be essentially as efficient as the MLE [5,6]. We derive the MMLEs of the parameters in (1.1) and show that they are considerably more efficient than the least squares estimators (LSEs). We also show that the MMLEs are robust to plausible deviations from an assumed distribution. This paper should be read in conjunction with [3] who give solutions when the autoregressive coefficient  $\phi$  is known to be zero, in which case (1.1) reduces to a simple multiple linear regression model.

## 2. Symmetric family

Assume that the iid errors  $e_i$  ( $1 \leq i \leq n$ ) have the distribution

$$f(e) = \frac{1}{\sigma \sqrt{k\beta} (1/2, p-1/2)} \left(1 + \frac{e^2}{k\sigma^2}\right)^{-p}, \quad -\infty < e < \infty; \quad (2.1)$$

$k = 2p - 3$  ( $p \geq 2$ );  $E(e) = 0$  and  $V(e) = \sigma^2$ . It may be noted that  $t = \sqrt{(v/k)} (e/\sigma)$  has Student's  $t$  distribution with  $v = 2p - 1$  degrees of freedom. The family (2.1) represents symmetric distributions with kurtosis greater than or equal to 3 and has been central to much statistical theory and practice [7,3,8–10]. Let

$$(y_i, x_{1,i}, \dots, x_{q,i}) \quad (1 \leq i \leq n) \quad (2.2)$$

be a random sample of size  $n$ . Conditional to the initial value,  $y_0$ , the likelihood function is

$$L\alpha \left(\frac{1}{\sigma}\right)^n \prod_{i=1}^n \left(1 + \frac{z_i^2}{k}\right)^{-p}; \quad (2.3)$$

$$z_i = e_i/\sigma \quad \text{and} \quad e_i = y_i - \phi y_{i-1} - \mu - \delta_1 (x_{1,i} - \phi x_{1,i-1}) - \dots - \delta_q (x_{q,i} - \phi x_{q,i-1}). \quad (2.4)$$

The likelihood equations  $\partial \ln L / \partial \mu = 0$ ,  $\partial \ln L / \partial \delta_j = 0$ , etc., for estimating  $\mu$ ,  $\delta_j$  ( $1 \leq j \leq q$ ),  $\phi$  and  $\sigma$  involve the functions

$$g(z_i) = z_i / \{1 + (1/k) z_i^2\} \quad (1 \leq i \leq n). \quad (2.5)$$

Solving them is enormously problematic realizing that they are nonlinear due to the parameters  $\delta_j \phi$  ( $1 \leq j \leq q$ ). The likelihood equations also have convergence problems if the sample contains outliers [11]. Therefore, we work out MMLEs as follows.

**Modified likelihood estimators:** Let  $z_{(i)} = e_{(i)}/\sigma$  be the ordered variates obtained by ordering  $e_i$  ( $1 \leq i \leq n$ ) in ascending order of magnitude;  $z_{(1)} \leq z_{(2)} \leq \dots \leq z_{(n)}$ . The first step is to express the likelihood equations in terms of  $z_{(i)}$ . This is accomplished simply by replacing  $z_i$  by  $z_{(i)}$ . The second step is to linearize the functions  $g(z_{(i)})$ :

$$g(z_{(i)}) \cong \alpha_i + \beta_i z_{(i)}, \quad (1 \leq i \leq n). \quad (2.6)$$

The coefficients  $\alpha_i$  and  $\beta_i$  are obtained from the first two terms of a Taylor series expansion of  $g(z_{(i)})$  around  $t_{(i)} = E(z_{(i)})$ . That gives

$$\alpha_i = \frac{(2/k) t_{(i)}^3}{[1 + (1/k) t_{(i)}^2]^2} \quad \text{and} \quad \beta_i = \frac{1 - (1/k) t_{(i)}^2}{[1 + (1/k) t_{(i)}^2]^2}, \quad 1 \leq i \leq n. \quad (2.7)$$

The values of  $t_{(i)}$  are obtained from the equation

$$\frac{1}{\sqrt{k\beta} (1/2, p-1/2)} \int_{-\infty}^{t_{(i)}} \left(1 + \frac{z^2}{k}\right)^{-p} dz = \frac{i}{n+1}, \quad 1 \leq i \leq n. \quad (2.8)$$

Modified likelihood equations are obtained by replacing  $g(z_{(i)})$  in the likelihood equations by linear approximations (2.6). Although the algebra involved in solving them is formidable, the solutions (MMLEs) have beautiful closed forms. Writing  $w_i = y_{[i]} - \hat{\delta}_1 x_{1,[i]} - \hat{\delta}_2 x_{2,[i]} - \dots - \hat{\delta}_q x_{q,[i]}$  where  $(y_{[i]}, y_{[i]-1}, x_{1,[i]}, x_{1,[i]-1}, \dots, x_{q,[i]}, x_{q,[i]-1})$  is that observation  $(y_i, y_{i-1}, x_{1,i}, x_{1,i-1}, \dots, x_{q,i}, x_{q,i-1})$  which is associated with the  $i$ th ordered residual  $e_{(i)}$  ( $1 \leq i \leq n$ ),

$$\hat{\mu} = \bar{v}_{[\cdot]} - \sum_{j=1}^q \hat{\delta}_j \bar{u}_{j[\cdot]}, \quad \hat{\delta} = A^{-1} (G + H \hat{\sigma}), \quad \hat{\phi} = K + D \hat{\sigma}, \quad (2.9)$$

$$\text{and } \hat{\sigma} = (B + \sqrt{B^2 + 4nC}) / 2n; \quad (2.10)$$

$$\hat{\delta} = \begin{bmatrix} \hat{\delta}_1 \\ \hat{\delta}_2 \\ \vdots \\ \hat{\delta}_q \end{bmatrix}, \quad G = \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_q \end{bmatrix}, \quad G_j = \sum_{i=1}^n \beta_i (u_{j[i]} - \bar{u}_{j[.]}) v_{[i]} / \sum_{i=1}^n \beta_i (u_{j[i]} - \bar{u}_{j[.]})^2,$$

$$H = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_q \end{bmatrix}, \quad H_j = \sum_{i=1}^n \alpha_i (u_{j[i]} - \bar{u}_{j[.]}) / \sum_{i=1}^n \beta_i (u_{j[i]} - \bar{u}_{j[.]})^2,$$

$$A = \begin{bmatrix} 1 & A_{12} & A_{13} & \cdots & A_{1q} \\ A_{21} & 1 & A_{23} & \cdots & A_{2q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{q1} & A_{q2} & A_{q3} & \cdots & 1 \end{bmatrix}, \quad A_{jk} = \frac{\sum_{i=1}^n \beta_i (u_{j[i]} - \bar{u}_{j[.]}) u_{k[i]}}{\sum_{i=1}^n \beta_i (u_{j[i]} - \bar{u}_{j[.]})^2}$$

$$K = \sum_{i=1}^n \beta_i (w_{i-1} - \bar{w}_{\cdot}) w_i / \sum_{i=1}^n \beta_i (w_{i-1} - \bar{w}_{\cdot})^2,$$

$$D = \sum_{i=1}^n \alpha_i (w_{i-1} - \bar{w}_{\cdot}) / \sum_{i=1}^n \beta_i (w_{i-1} - \bar{w}_{\cdot})^2, \quad \bar{w}_{\cdot} = (1/m) \sum_{i=1}^n \beta_i w_{i-1} \quad \left( m = \sum_{i=1}^n \beta_i \right),$$

$$B = \frac{2p}{k} \sum_{i=1}^n \alpha_i [(v_{[i]} - \bar{v}_{[.]}) - U'A^{-1}G] \quad \text{and} \quad C = \frac{2p}{k} \sum_{i=1}^n \beta_i [(v_{[i]} - \bar{v}_{[.]}) - U'A^{-1}G]^2; \quad (2.11)$$

$$U = \begin{bmatrix} u_{1[i]} - \bar{u}_{1[.]} \\ u_{2[i]} - \bar{u}_{2[.]} \\ \vdots \\ u_{q[i]} - \bar{u}_{q[.]} \end{bmatrix}.$$

In the equations above,

$$v_{[i]} = y_{[i]} - \hat{\phi} y_{[i]-1}, \quad u_{j[i]} = x_{j,[i]} - \hat{\phi} x_{j,[i]-1}, \quad (2.12)$$

$$\bar{v}_{[.]} = \frac{1}{m} \sum_{i=1}^n \beta_i v_{[i]} \quad \text{and} \quad \bar{u}_{j[.]} = \frac{1}{m} \sum_{i=1}^n \beta_i u_{j[i]}.$$

The MMLEs (2.9) and (2.10) are natural extensions of those given in [3, p. 2450]. Linear approximations like (2.6) give very accurate values [12, p. 155] and, consequently, MMLEs are numerically the same (almost) as MLEs; see [13, p. 101–106]. Smith et al. [14] considered quadratic approximations to functions like  $g(z_{(i)})$  and noticed no worthwhile improvements in the efficiencies. They concluded that quadratic approximations only make the algebra very cumbersome.

**Computations:** To initialize ordering of  $e_{(i)}$  ( $1 \leq i \leq n$ ), we write  $d_j = -\delta_j \phi$  ( $1 \leq j \leq q$ ) and calculate the initial estimates by minimizing

$$\sum_{i=1}^n a_i^2 = \sum_{i=1}^n (y_i - \phi y_{i-1} - \delta_1 x_{1,i} - \cdots - \delta_q x_{q,i} - d_1 x_{1,i-1} - \cdots - d_q x_{q,i-1})^2. \quad (2.13)$$

Since  $\mu$  is a constant and  $\sigma > 0$ ,  $z_{(i)} = e_{(i)}/\sigma$  are determined by the ordered  $a_{(i)}$  ( $1 \leq i \leq n$ ) values. Denote the solutions, obtained by minimizing (2.13), by  $\hat{\phi}_0$ ,  $\hat{\delta}_{j0}$  and  $\hat{d}_{j0}$  ( $1 \leq j \leq q$ ). Initially, therefore,

$$a_{(i)} = y_{[i]} - \hat{\phi}_0 y_{[i]-1} - \hat{\delta}_{10} x_{1,[i]} - \cdots - \hat{\delta}_{q0} x_{q,[i]} - \hat{d}_{10} x_{1,[i]-1} - \cdots - \hat{d}_{q0} x_{q,[i]-1}, \quad (1 \leq i \leq n).$$

Using the concomitants  $(y_{[i]}, x_{1,[i]}, \dots, x_{q,[i]})$  ( $1 \leq i \leq n$ ), the MMLEs  $\hat{\delta}_j$  ( $1 \leq j \leq q$ ) and  $\hat{\sigma}$  are calculated from the equations above with  $\phi = \hat{\phi}_0$ . The MMLE  $\hat{\phi}$  is then calculated from (2.9). A second iteration is carried out with  $\hat{\phi}_0$ ,  $\hat{\delta}_{j0}$  and  $\hat{d}_{j0}$  replaced by  $\hat{\phi}$ ,  $\hat{\delta}_j$  and  $-\hat{\phi}\hat{\delta}_j$ , respectively. One more iteration is carried out to obtain the final estimates. In all our computations only two iterations were required for the estimates to stabilize sufficiently enough. This is because only the relative magnitudes of  $a_i$  (not necessarily their true values) are required for locating the concomitant observations. The MMLE  $\hat{\mu}$  is then computed from (2.9).

**Remark.** The LSEs (least squares estimators) are calculated exactly the same way as the MMLEs with  $\alpha_i$  and  $\beta_i$  equated to 0 and 1, respectively, and  $2p/k$  in (2.11) equated to 1. We denote the LSE by  $\tilde{\mu}$ ,  $\tilde{\delta}_j$  ( $1 \leq j \leq q$ ),  $\tilde{\phi}$  and  $\tilde{\sigma}$ .

**Comment:** For small  $p (< 3.5)$ ,  $\beta_1$  (and few other coefficients  $\beta_i$ ) can be negative as a result of which  $\hat{\sigma}$  can cease to be real and positive. To rectify this situation, if  $\beta_1 < 0$ ,  $\alpha_i$  and  $\beta_i$  are equated to [7, p. 409]  $\alpha_i^* = (1/k) t_{(i)}^3 / \left\{ 1 + (1/k) t_{(i)}^2 \right\}^2$  and  $\beta_i^* = 1 / \left\{ 1 + (1/k) t_{(i)}^2 \right\}^2$ , respectively, and the MMLs calculated. The resulting estimator  $\hat{\sigma}$  is always real and positive. This does not alter the asymptotic properties of the MMLs since the function  $g(z)$  is bounded and so are  $\alpha_i$  and  $\beta_i$ . Consequently,  $z_{(i)} - t_{(i)} \cong 0$  and

$$\alpha_i + \beta_i z_{(i)} \cong \alpha_i^* + \beta_i^* z_{(i)}. \quad (2.14)$$

**Asymptotic covariance matrix:** Since MMLs are asymptotically equivalent to MLEs [4, Appendix A], their variances and covariances are given by  $I^{-1}$ ,  $I$  being the Fisher information matrix. To derive  $I$ , we need  $E(y_i)$  and  $V(y_i)$ . The variance  $V(y_i) = \sigma^2 / (1 - \phi^2)$  and if  $E(y_0) = 0$ ,

$$E(y_i) = \mu \sum_{j=1}^n \phi^{j-1} + \sum_{j=1}^n \delta_j x_{j,i} - \phi^i \sum_{j=1}^n \delta_j x_{j,0} \quad (1 \leq i \leq n). \quad (2.15)$$

Since  $\delta_j$  are not necessarily small, the last term on the right hand side cannot be ignored to make the derivation of  $I$  manageable. Therefore, we recommend the use of the sample information matrix [15]. A sample information matrix consists of the elements  $-\partial^2 \ln L / \partial \theta_i \partial \theta_j$  ( $i, j = 1, 2, \dots, k; k \geq 1$ ) evaluated at  $\theta_i = \hat{\theta}_i$  and  $\theta_j = \hat{\theta}_j$ . The inverse of the information matrix gives accurate approximations to the variances and covariances for large  $n$ ; see the Appendix. See also [16, Table 1].

**Remark.** In calculating the MMLs,  $\alpha_i^*$  and  $\beta_i^*$  should be used only if  $\beta_1 < 0$ . Otherwise, a little loss in efficiencies occurs unless  $n$  is large ( $n > 100$ ); see [10, p. 1731].

## 2.1. Simulations for the symmetric family

The LSEs are used very extensively in practice; see, for example, [17–20]. We will show that they are considerably less efficient than the MMLs for the family (2.1). Given in Table 1 are the simulated values, based on [100, 000/ $n$ ] (integer value) Monte Carlo runs, of the means of the estimators of  $\phi$  and  $\sigma$ , the variances of the MMLs, and the relative efficiencies

$$RE = 100(\text{variance of the MML} / \text{variance of the LSE}) \quad (2.1.1)$$

of the LSE. It may be noted that increasing the number of simulation runs does not change the values in any substantial way; see also [7,9]. The means of the estimators of  $\mu$  and  $\delta_j$  are not reported since the bias in them was found to be negligible for all sample sizes. We give values only for  $q = 2$  and  $n = 50$  for conciseness. Without loss of generality, we take  $\mu = 0$ ,  $\delta_j = 1$  and  $\sigma = 1$ . It can be seen that the MMLs are enormously more efficient than the LSEs, and the relative efficiencies are essentially the same for all values of  $\phi$ . Incidentally,  $y_0$  was taken to be equal to  $e_0 / \sqrt{1 - \phi^2}$  where  $e_0$  is a random error having the same distribution as that of  $e_i$  ( $1 \leq i \leq n$ ). Thus,  $V(y_0) = \sigma^2 / (1 - \phi^2)$ . This is, in fact, Model II of [21]. The design values  $x_{ij}$ 's were generated from a uniform distribution and divided by  $\sqrt{1 - \phi^2}$  as in [22,23]. This gives a wider spread to the design values realizing that  $V(y_i) = \sigma^2 / (1 - \phi^2)$ . We do not reproduce the values for negative  $\phi$  since they are essentially the same as for the corresponding positive values of  $\phi$  due to symmetry of the family (2.1).

A disconcerting feature of the LSE is that their relative efficiencies decrease as the sample size  $n$  increases. For  $n = 100$  and  $\phi = 0.5$ , for example, we have the following simulated values. These may be compared with the values given in Table 1.

Means, variances and relative efficiencies ( $q = 2$ ); $n = 100, \phi = 0.5$ .												
Mean		$n \times \text{Var}$				Mean		$n \times \text{Var}$				
$\phi$	$\sigma$	$\mu$	$\delta_1$	$\phi$	$\sigma$	$\phi$	$\sigma$	$\mu$	$\delta_1$	$\phi$	$\sigma$	
$p = 2$						$p = 2.5$						
MML	0.487	1.076	0.662	0.305	0.516	3.304	0.484	1.042	0.869	0.424	0.620	2.049
LS	0.482	0.943	59	54	70	47	0.479	0.967	78	72	81	78
$p = 3.5$						$p = 5$						
MML	0.482	0.975	1.043	0.500	0.710	0.932	0.481	0.986	1.029	0.504	0.751	0.666
LS	0.480	1.018	91	88	91	86	0.480	0.977	93	93	95	87

The values for  $\delta_2$  are not reproduced for the same reason as in Table 1.

**Robustness:** Since deviations from an assumed distribution are very common, one cannot feel comfortable with assuming a particular distribution and believing it to be exactly correct [13, Preface]; see also [24, Preface]. It is, therefore, very important for a statistical procedure to be robust to plausible deviations from an assumed distribution. We now illustrate that, compared to the LSE, the MMLs are remarkably robust, i.e., they are fully efficient (or nearly so) for an assumed distribution and maintain high efficiency for plausible alternatives. Consider, for example, the situation when the underlying distribution is (2.1) with  $p = 3.5$  (a scaled Student's  $t$  distribution with  $\nu = 6$  degrees of freedom); we denote (2.1) by  $f(p, \sigma)$ . As plausible alternatives, we consider the following.

**Table 1**

Means and variances of the MMLEs, and the relative efficiencies of the LSE;  $\mu = 0, \delta_1 = \delta_2 = 1 (q = 2), \sigma = 1$ , sample size  $n = 50$ .

		$\phi = 0.5$						$\phi = 0.9$					
		Mean		$n \times \text{Var}$				Mean		$n \times \text{Var}$			
		$\phi$	$\sigma$	$\mu$	$\delta_1$	$\phi$	$\sigma$	$\phi$	$\sigma$	$\mu$	$\delta_1$	$\phi$	$\sigma$
$p = 2$	MML	0.472	1.103	0.775	0.374	0.593	3.210	0.866	1.099	1.819	0.077	0.174	3.766
	LS	0.464	0.900	65	63	77	81	0.855	0.896	63	61	71	81
$p = 2.5$	MML	0.466	1.058	1.020	0.306	0.688	1.779	0.841	1.057	2.577	0.082	0.320	1.742
	LS	0.460	0.941	81	77	86	86	0.832	0.938	80	77	86	85
$p = 3.5$	MML	0.468	1.019	1.084	0.428	0.697	0.983	0.827	0.946	2.970	0.096	0.374	1.002
	LS	0.466	0.951	93	90	93	90	0.831	1.014	92	91	93	90
$p = 5$	MML	0.465	0.968	1.136	0.560	0.801	0.681	0.842	0.969	2.932	0.093	0.311	0.700
	LS	0.463	0.951	94	95	97	92	0.839	0.951	93	94	96	92

The results for the MML and the LS estimators of  $\delta_2$  are essentially the same as for  $\delta_1$  and are not, therefore, reported.

**Table 2**

Means and variances of the MMLEs, and the relative efficiencies of the LSE ( $q = 2$ );  $\mu = 0, \delta_1 = \delta_2 = 1 (q = 2), \phi = 0.5, \sigma = 1$ , sample size  $n = 50$ .

Model			Mean <sup>a</sup>				$n \times \text{Var}^a$			
			$\mu$	$\delta_1$	$\phi$	$\sigma$	$\mu$	$\delta_1$	$\phi$	$\sigma$
1	MML	$\tau = 1$	0.000	1.000	0.487	1.076	0.662	0.305	0.516	4.219
	LS		0.000	1.000	0.482	0.943	59	54	70	47
2	MML	$\tau = 1$	−0.001	1.000	0.484	1.042	0.869	0.424	0.620	2.049
	LS		−0.001	0.999	0.479	0.967	78	72	81	60
3	MML	$\tau = 1$	0.000	1.000	0.482	1.018	1.043	0.500	0.710	0.932
	LS		0.000	1.000	0.480	0.975	91	88	91	86
4	MML	$\tau = 1$	0.006	0.998	0.481	1.044	1.268	0.627	0.830	0.651
	LS		0.006	0.999	0.463	0.954	101	105	104	101
5	MML	$\tau = 1.581$	−0.007	1.001	0.462	1.545	2.243	1.683	1.518	6.614
	LS		−0.007	1.003	0.440	1.516	72	69	82	82
6	MML	$\tau = 1.581$	0.001	1.008	0.469	1.473	2.180	0.911	0.546	7.103
	LS		0.000	1.009	0.463	1.441	73	70	81	84
7	MML	$\tau = 0.953$	0.005	1.000	0.464	1.000	1.079	0.733	0.716	0.891
	LS		0.005	1.000	0.462	0.930	94	93	95	99

<sup>a</sup> The values for  $\delta_2$  are not given for the same reason as in Table 1.

Misspecification of the distribution:

$$(1) p = 2, \quad (2) p = 2.5, \quad (3) p = 5, \quad (4) p = \infty (\text{normal distribution}). \quad (2.1.2)$$

Outlier model:

$$(5) (n - r) \text{ observations come from } f(3.5, \sigma) \text{ and } r \text{ (we do not know which) come from } f(3.5, 4\sigma), \\ r = [0.5 + 0.1n] \text{ (integer value)}. \quad (2.1.3)$$

Mixture model:

$$(6) 0.90f(3.5, \sigma) + 0.10f(3.5, 4\sigma). \quad (2.1.4)$$

Contamination model:

$$(7) 0.90f(3.5, \sigma) + 0.10\text{Uniform}(-0.5, 0.5). \quad (2.1.5)$$

We simulated the means, the variances of the MMLEs, and the relative efficiencies of the LSE. The values are given in Table 2. It may be noted that both  $\hat{\sigma}$  and  $\tilde{\sigma}$  are estimating  $\tau\sigma$ ;  $\tau$  is, in fact, equal to the square root of the variance of the sample model to the variance of the population model. It has absolutely no role to play in the computation of the estimators. Its values are given in Table 2 only for determination of the bias in  $\hat{\sigma}$  and  $\tilde{\sigma}$ . It can be seen that the MMLEs are overall considerably more efficient than the LSE. For model 4, of course, the MMLEs are a little less efficient than the LSE as expected.

**Remark.** We repeated the simulations above with other designs, e.g.,  $x_{ij}$  generated from normal  $N(0, \sigma^2)$  as in [25]. The results were found to be essentially the same as those in Table 2.

**Comment:** The robustness of the MMLEs is due to the fact that the  $\beta_i (1 \leq i \leq n)$  coefficients increase until the middle value and then decrease in a symmetric fashion. Thus, the extreme residuals automatically receive small weights. That depletes the effect of long-tails and outliers.

### 3. Generalized Logistic family

Consider the situation when the iid errors in (1.1) have one of the distributions in the Generalized Logistic family

$$f(e) = \frac{b}{\sigma} \exp(-e/\sigma) / \{1 + \exp(-e/\sigma)\}^{b+1}, \quad -\infty < e < \infty. \quad (3.1)$$

For  $b < 1$  and  $b > 1$ , (3.1) represents negatively and positively skewed distributions, respectively. For  $b = 1$ , (3.1) is the well-known logistic distribution. The family (3.1) has played a pivotal role in modeling data in numerous areas of application [22, 26, 3, 16, 5, 27]. The likelihood equations for estimating  $\mu$ ,  $\delta_j$  ( $1 \leq j \leq q$ ),  $\phi$  and  $\sigma$  involve the functions

$$g(z_i) = 1 / \{1 + \exp(z_i)\} \quad (1 \leq i \leq n), \quad (3.2)$$

and are almost impossible to solve. The modified likelihood equations are obtained by linearizing  $g(z_{(i)})$ , exactly along the same lines as in Section 2. Their solutions are the following MMLs:

$$\hat{\mu} = \bar{v}_{[.]} - \sum_{j=1}^q \hat{\delta}_j \bar{u}_{j[.]} + (\Delta/m) \hat{\sigma}; \quad (3.3)$$

$\hat{\delta}_j$ ,  $\hat{\phi}$  and  $\hat{\sigma}$  have exactly the same expressions as those in (2.9)–(2.12) with the following changes.

In the expression for  $H_j$ ,  $\alpha_i$  is replaced by

$$\Delta_i = (b+1)^{-1} - \alpha_i \quad (1 \leq i \leq n), \quad \text{and} \\ D = \sum_{i=1}^n \Delta_i (w_{i-1} - \bar{w}) / \sum_{i=1}^n \beta_i (w_{i-1} - \bar{w})^2, \quad \Delta = \sum_{i=1}^n \Delta_i. \quad (3.4)$$

In the expressions for B and C, the multiplying constant  $2p/k$  is replaced by  $b+1$ . The coefficients  $\alpha_i$  and  $\beta_i$  in (2.7) are replaced by

$$\alpha_i = [1 + \exp(t_{(i)})]^{-1} + \beta_i t_{(i)} \quad \text{and} \quad \beta_i = \exp(t_{(i)}) / [1 + \exp(t_{(i)})]^2; \quad (3.5) \\ t_{(i)} = -\ln \left[ \left( \frac{i}{n+1} \right)^{-1/b} - 1 \right], \quad (1 \leq i \leq n).$$

It is interesting to see that the MMLs have exactly the same expressions as in (2.9)–(2.12) in spite of the fact that the families of distributions (2.1) and (3.1) are very different from one another. Here, the MMLE  $\hat{\sigma}$  is always real and positive since the coefficients  $\beta_i$  in (3.5) are positive for all  $i = 1, 2, \dots, n$ .

### 3.1. Simulations for the Generalized Logistic family

To evaluate the relative efficiencies of the LSEs, we simulated the means and the variances of the MMLs and the LSEs for  $q = 1, 2$  and 4. We give the values in Table 3 only for  $q = 2$  and  $n = 50$ , for conciseness. The values for  $q \neq 2$  are exactly similar. Since the MMLs and the LSEs of  $\delta_1$  and  $\delta_2$  have negligible bias, we do not reproduce their means.

From the relative efficiencies given in Table 3, it can be seen that the MMLs are considerably more efficient than the LSE. Both the MML and the LSE of  $\mu$  have some bias for  $b \neq 1$  due to the fact that (3.1) is a family of skew distributions. However, the bias decreases as  $n$  increases. For  $n = 100$ , for example, we have the following values;  $\mu = 0$ ,  $\delta_1 = \delta_2 = 1$  ( $q = 2$ ) and  $\sigma = 1$  without loss of generality.

Simulated means, variances and relative efficiencies; $n = 100, b = 0.5$ .								
$\phi$		Mean			$N \times \text{Var}$			
		$\mu$	$\phi$	$\sigma$	$\mu$	$\delta_1$	$\phi$	$\sigma$
-0.5	MML	-0.029	-0.501	0.990	5.664	2.681	0.553	0.787
	LS	-0.037	-0.504	0.974	94	78	86	74
0.5	MML	-0.086	0.484	0.996	10.88	3.102	0.671	0.765
	LS	-0.098	0.478	0.977	89	84	85	75
0.9	MML	-0.345	0.875	0.993	38.67	0.460	0.189	0.764
	LS	-0.422	0.869	0.978	77	76	76	75

To have a better perspective of the bias in the MML and the LSE of  $\mu$ , their simulated means given here and in Table 3 should be multiplied by  $\sqrt{1 - \phi^2}$  since  $V(y_i) = \sigma^2 / (1 - \phi^2)$ .

**Comment:** The MMLs are, as compared to the LSE, robust because the  $\beta_i$  ( $1 \leq i \leq n$ ) coefficients decrease in the direction of the long-tail. Thus, the effect of the long-tail is automatically depleted; see also [3].

## 4. Short-tailed symmetric family

A coherent family of STS distributions is given by [27]

$$f(e) = \frac{C_1}{\sigma} \left\{ 1 + \frac{\lambda}{2r} \left( \frac{e}{\sigma} \right)^2 \right\}^r \frac{\exp(-e^2 / 2\sigma^2)}{\sqrt{2\pi}}, \quad -\infty < e < \infty; \quad (4.1)$$

**Table 3**Means and variances of the MMLs, and the relative efficiencies of the LSE;  $\mu = 0, \delta_1 = \delta_2 = 1$  ( $q = 2$ ) and  $\sigma = 1$ . Sample size  $n = 50$ .

$b$	$\phi$		Mean <sup>a</sup>			$n \times \text{Var}^a$			
			$\mu$	$\phi$	$\sigma$	$\mu$	$\delta_1$	$\phi$	$\sigma$
0.5	−0.5	MML	−0.057	−0.505	0.976	6.629	2.767	0.667	0.759
		LS	−0.063	−0.509	0.951	93	79	84	82
	0.5	MML	−0.163	0.468	0.976	12.75	3.600	0.686	0.753
		LS	−0.189	0.459	0.945	88	78	82	77
	0.9	MML	−0.723	0.847	0.979	56.23	0.567	0.290	0.739
		LS	−0.882	0.835	0.954	83	78	84	77
2	−0.5	MML	0.058	−0.500	0.976	2.501	1.749	0.663	0.666
		LS	0.054	−0.503	0.954	102	89	91	84
	0.5	MML	1.240	0.461	0.970	6.538	1.240	0.800	0.652
		LS	1.420	0.456	0.944	99	87	92	83
	0.9	MML	0.230	0.857	0.974	19.91	0.230	0.234	0.710
		LS	0.252	0.853	0.947	96	91	90	88
4	−0.5	MML	0.165	−0.498	0.982	4.033	1.273	0.678	0.600
		LS	0.099	−0.500	0.947	101	85	87	73
	0.5	MML	0.281	0.469	0.983	13.69	0.983	0.715	0.618
		LS	0.220	0.463	0.952	97	77	87	71
	0.9	MML	0.427	0.880	0.970	24.58	0.169	0.091	0.618
		LS	0.471	0.876	0.949	82	81	77	73
6	−0.5	MML	0.317	−0.507	0.993	5.332	0.911	0.621	0.655
		LS	0.129	−0.508	0.953	95	76	90	73
	0.5	MML	0.313	0.473	0.986	18.43	1.061	0.671	0.607
		LS	0.258	0.464	0.959	92	79	86	65
	0.9	MML	0.369	0.888	0.962	21.55	0.130	0.051	0.628
		LS	0.374	0.886	0.949	78	77	76	69

<sup>a</sup> The values for  $\delta_2$  are essentially the same as for  $\delta_1$ . $\lambda = r/(r - d)$ ,  $d < r$  and  $r$  is an integer, and

$$C_1 = 1 / \left\{ \sum_{j=0}^r \binom{r}{j} \left( \frac{\lambda}{2r} \right)^j \frac{(2j)!}{2^j(j)!} \right\}.$$

In practice, it suffices to take  $r = 4$ . The reason is that (4.1) represents a broad range of STS distributions for  $r = 4$ ; see [27, p. 1023]; (4.1) represents symmetric distributions with kurtosis less than or equal to 3 and is very useful in modeling data in several situations [28,24].

The values of  $C_1$  and the moments  $E(z^{2j})$  are obtained from the equation

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^{2j} \exp\left(-\frac{1}{2}t^2\right) dt = \frac{(2j)!}{2^j(j)!}. \quad (4.2)$$

Note that no distribution can have kurtosis less than 1 [29].

With (4.1) as the distribution of  $e_i$  ( $1 \leq i \leq n$ ), the maximum likelihood equations are expressions in terms of the intractable functions  $g(z_i) = z_i/\{1 + (\lambda/2r)z_i^2\}$ . The MLEs of the parameters in the model (1.1) are, therefore, elusive. The MMLs can, however, be obtained exactly the same way as in Section 2. They are

$$\hat{\delta} = A^{-1}(G - H\lambda\hat{\sigma}), \quad \hat{\phi} = K - D\lambda\hat{\sigma} \quad \text{and} \quad \hat{\sigma} = \left( -\lambda B + \sqrt{(\lambda B)^2 + 4nC} \right) / 2n; \quad (4.3)$$

$\hat{\mu}$  has exactly the same expression as in (2.9). The expressions for A, B, C, D, G, H and K are exactly the same as those in (2.11) with  $2p/k$  replaced by 1 and the coefficients  $\alpha_i$  and  $\beta_i$  replaced by

$$\alpha_i = (\lambda/r)t_{(i)}^3 / \{1 + (\lambda/2r)t_{(i)}^2\}^2 \quad \text{and} \quad \beta_i = 1 - \lambda\gamma_i; \quad (4.4)$$

$$\gamma_i = \{1 - (\lambda/2r)t_{(i)}^2\} / \{1 + (\lambda/2r)t_{(i)}^2\}^2 \quad (1 \leq i \leq n), \quad \text{for } \lambda \leq 1.$$

The values of  $t_{(i)}$  are obtained from the equation ( $z = e/\sigma$ )

$$\int_{-\infty}^{t_{(i)}} f(z) dz = \frac{i}{n+1}, \quad 1 \leq i \leq n. \quad (4.5)$$

A simple algorithm to calculate  $t_{(i)}$  is available with the authors.

For  $\lambda \leq 1$ , the coefficients  $\beta_i$  are all positive. Hence,  $\hat{\sigma}$  is real and positive. For  $\lambda > 1$ , however, a few coefficients in the middle can be negative as a result of which  $\hat{\sigma}$  can cease to be real. To rectify this situation,  $\alpha_i$  and  $\beta_i$  in (4.4) are replaced by  $\alpha_i^*$  and  $\beta_i^*$ , respectively:



**Table 4**Variances of the MMLEs and the relative efficiencies of the LSE;  $\mu = 0$ ,  $\delta_1 = \delta_2 = 1$  ( $q = 2$ ),  $\phi = 0.5$ ,  $\sigma = 1$  and  $n = 50$ .

$r = 4$ Parameter	$d = -0.5$		$d = 1$		$d = 2$		$d = 3$	
	$n \times \text{Var}$	RE	$n \times \text{Var}$	RE	$n \times \text{Var}$	RE	$n \times \text{Var}$	RE
$\mu$	2.715	90	3.255	78	3.730	67	7.475	94
$\delta_1$	0.995	91	1.045	74	1.575	49	2.285	52
$\phi$	0.785	93	0.640	74	0.480	52	0.435	52
$\sigma$	0.335	92	0.265	88	0.195	89	0.125	96

The values for  $\delta_2$  are essentially the same as for  $\delta_1$ .

$$\alpha_i^* = \{(\lambda/r)t_{(i)}^3 + (1 - 1/\lambda)t_{(i)}\} / \{1 + (\lambda/2r)t_{(i)}^2\}^2 \quad \text{and} \quad \beta_i^* = 1 - \lambda\gamma_i^*; \quad (4.6)$$

$$\gamma_i^* = \{(1/\lambda) - (\lambda/2r)t_{(i)}^2\} / \{1 + (\lambda/2r)t_{(i)}^2\}^2 \quad (1 \leq i \leq n), \text{ for } \lambda > 1.$$

Note that  $\beta_i^*$  are all positive for  $\lambda > 1$ . For  $\lambda = 1$ ,  $\alpha_i = \alpha_i^*$  and  $\beta_i = \beta_i^*$  ( $1 \leq i \leq n$ ). Asymptotically,  $z_{(i)} - t_{(i)} \cong 0$  and

$$g\{z_{(i)}\} \cong \alpha_i + \beta_i z_{(i)} \quad (\lambda \leq 1) \quad (4.7)$$

$$\cong \alpha_i^* + \beta_i^* z_{(i)} \quad (\lambda \geq 1), \quad (1 \leq i \leq n). \quad (4.8)$$

#### 4.1. Simulations for the short-tailed symmetric family

We simulated the means and variances of the MMLEs and LSEs for  $n = 20, 30, 50$  and  $100$ . The bias in all the estimators is negligible. The MMLEs are, however, considerably more efficient than the LSE. For  $n = 50$ , the variances of the MMLEs and the relative efficiencies of the LSEs are given in Table 4. The design values  $x_{ji}$  were generated from a uniform distribution as in Table 1.

The MMLEs are robust to plausible deviations from an assumed STS distribution and to inliers (discrepant observations in the middle of the sample). Two inlier models are proposed in [27,28]. The robustness of the MMLE is due to the fact that the coefficients  $\beta_i$  and  $\beta_i^*$  ( $1 \leq i \leq n$ ) in (4.4)–(4.6) decrease until the middle value and then increase again in a symmetric fashion. Thus, the discrepant observations in the middle automatically receive small weights and their influence is depleted.

### 5. Determination of shape parameters

In practice, the shape parameters  $p$ ,  $b$  and  $d$  in (2.1), (3.1) and (4.1), respectively, may not be known. Their plausible values can be determined as follows [3, Examples 1–3], [7, Example 2].

We first calculate the LSE and the estimated deviants

$$\tilde{w}_i = y_i - \tilde{\phi}y_{i-1} - \sum_{j=1}^q \tilde{\delta}_j (x_i - \tilde{\phi}x_{i-1}), \quad (1 \leq i \leq n). \quad (5.1)$$

We plot the order statistics  $\tilde{w}_{(i)}$  against the quantiles  $Q_i$  of a standard normal distribution:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Q_i} \exp\left(-\frac{1}{2}t^2\right) dt = \frac{i}{n+1}, \quad 1 \leq i \leq n. \quad (5.2)$$

The plot (called  $Q$ – $Q$  plot) gives useful information about the nature of the underlying distribution; see specifically [30, p. 16]. See also [5,3]. Suppose that the distribution is one in the family (2.1). We now calculate the MMLEs for a series of values of  $p$ . We calculate the corresponding values of

$$\ln \hat{L} = \ln L \text{ (evaluated at } \mu = \hat{\mu}, \delta_j = \hat{\delta}_j, \phi = \hat{\phi}, \sigma = \hat{\sigma}). \quad (5.3)$$

The value of  $p$  that yields  $\max(\ln \hat{L})$  is the chosen value. This procedure gives a value in close proximity (if not equal) to the true value of the shape parameter, at any rate for large  $n$ . The corresponding MMLEs have the desired optimality properties because of their intrinsic robustness to plausible deviations from the true distribution. See [9, Section 8].

**Example.** Bass and Clarke give data to determine the effect of advertising on sales and state that the effect of advertising in one period carries over to the next. The data consists of 36 pairs of observations  $(X, Y)$ , where  $Y$  represents sales and  $X$  the advertising. The data is reproduced from [26, p.83].

X:	15.0	16.0	18.0	27.0	21.0	49.0	21.0	22.0	28.0	36.0	40.0	3.0	21.0
Y:	12.0	20.5	21.0	15.5	15.3	23.5	24.5	21.3	23.5	28.0	24.0	15.5	17.3
X:	29.0	62.0	65.0	46.0	44.0	33.0	62.0	22.0	12.0	24.0	3.0	5.0	14.0
Y:	25.3	25.0	36.5	36.5	29.6	30.5	28.0	26.0	21.5	19.7	19.0	16.0	20.7
X:	36.0	40.0	49.0	7.0	52.0	65.0	17.0	5.0	17.0	1.0			
Y:	26.5	30.6	32.3	29.5	28.3	31.3	32.2	26.4	23.4	16.4			

Since an observation is influenced by the previous one (Eq. (1.1)) and the observation prior to the first pair (15.0, 12.0) is not available for a visual inspection, it is advisable to disregard the first pair. The second pair (16.0, 20.5) is taken to



be  $(x_0, y_0)$ . We have  $n = 34$  additional pairs of observations. We assume the model (1.1) and calculate the LSE of  $\delta_1$  and  $\phi$ ;  $q = 1$ :

$$\tilde{\delta}_1 = 0.0883 \quad \text{and} \quad \tilde{\phi} = 0.629.$$

We calculate the deviants

$$\tilde{w}_i = y_i - \tilde{\phi}y_{i-1} - \tilde{\delta}_1(x_i - \tilde{\phi}x_{i-1}), \quad (1 \leq i \leq 34).$$

We plot the ordered deviants  $\tilde{w}_{(i)}$  against the quantiles  $Q_i$  of the normal  $N(0, 1)$ . It indicates a short-tailed symmetric distribution, e.g., a member of the family (4.1). We now calculate the MMLEs for a series of values of  $d$  including  $d = 0$ . The results are given below:

$$\begin{array}{ccc} d = & -0.5 & 0 & 0.5 \\ (1/n) \ln \hat{L} = & -2.768 & -2.765 & -2.767. \end{array}$$

Thus,  $d = 0$  ( $r = 4$ ) in (4.1) is the most plausible value; the variance of the distribution is  $2.576\sigma^2$ . The LSEs and the MMLEs are

$$\begin{array}{l} \text{LSE: } \tilde{\delta} = 0.0883, \quad \tilde{\phi} = 0.629 \quad \text{and} \quad \tilde{\sigma} = 2.553; \\ \text{MMLE: } \hat{\delta} = 0.102, \quad \hat{\phi} = 0.639 \quad \text{and} \quad \hat{\sigma} = 2.506. \end{array}$$

The MMLEs indicate a stronger effect of advertising on sales than do the LSEs.

## 6. Conclusion

We have considered estimation of parameters in a multiple autoregressive model. Since MLEs (maximum likelihood estimators) are elusive, we have derived MMLEs (modified maximum likelihood estimators) of the unknown parameters. The latter are explicit functions of sample observations and are easy to compute. Using asymptotic mathematics we have shown that MMLEs are fully efficient (unbiased with minimum variances). Using simulations, we have shown that MMLEs are considerably more efficient than LSEs (least squares estimators) for all  $n$  (sample size). Another important problem is to test the null hypothesis  $H_0 : \phi = 0$  or  $\delta_j = 0$  ( $j = 1, 2, \dots, q$ ). If  $\phi = 0$ , model (1.1) reduces to a multiple linear regression model in which case the methodology developed in [3] becomes applicable. Testing  $H_0$ , however, is a difficult problem and will be the subject matter of a future paper.

## Acknowledgements

Thanks are due to the referees for very valuable comments. We are very grateful to Professor M.L. Tiku for his help and advice in the course of this research.

## Appendix

Consider the family of distributions (2.1). The second derivatives are:

$$\begin{aligned} -\frac{\partial^2 \ln L}{\partial \mu^2} &= \frac{2p}{k\sigma^2} \sum_{i=1}^n h(z_i), & -\frac{\partial^2 \ln L}{\partial \delta_j \partial \mu} &= \frac{2p}{k\sigma^2} \sum_{i=1}^n u_{ji} h(z_i) \quad (1 \leq j \leq q) \\ -\frac{\partial^2 \ln L}{\partial \phi \partial \mu} &= \frac{2p}{k\sigma^2} \sum_{i=1}^n w_{i-1} h(z_i), & -\frac{\partial^2 \ln L}{\partial \sigma \partial \mu} &= \frac{2p}{k\sigma^2} \left[ \sum_{i=1}^n g(z_i) + \sum_{i=1}^n z_i h(z_i) \right], \\ -\frac{\partial^2 \ln L}{\partial \delta_j^2} &= \frac{2p}{k\sigma^2} \sum_{i=1}^n u_{ji}^2 h(z_i), & (1 \leq j \leq q), \\ -\frac{\partial^2 \ln L}{\partial \delta_l \partial \delta_j} &= \frac{2p}{k\sigma^2} \sum_{i=1}^n u_{ji} u_{li} h(z_i), & (1 \leq j \leq q-1, j+1 \leq l \leq q) \\ -\frac{\partial^2 \ln L}{\partial \phi \partial \delta_j} &= \frac{2p}{k\sigma^2} \sum_{i=1}^n u_{ji} w_{i-1} h(z_i), & (1 \leq j \leq q) \\ -\frac{\partial^2 \ln L}{\partial \sigma \partial \delta_j} &= \frac{2p}{k\sigma^2} \left[ \sum_{i=1}^n u_{ji} g(z_i) + \sum_{i=1}^n u_{ji} z_i h(z_i) \right], & (1 \leq j \leq q) \\ -\frac{\partial^2 \ln L}{\partial \phi^2} &= \frac{2p}{k\sigma^2} \sum_{i=1}^n w_{i-1}^2 h(z_i) & -\frac{\partial^2 \ln L}{\partial \sigma \partial \phi} &= \frac{2p}{k\sigma^2} \left[ \sum_{i=1}^n w_{i-1} g(z_i) + \sum_{i=1}^n w_{i-1} z_i h(z_i) \right] \end{aligned}$$

**Table A.1**Values of  $n \times$  Variances obtained from the sample information matrix;  $\mu = 0$ ,  $\delta_1 = \delta_2 = 1$  ( $q = 2$ ),  $\sigma = 1$ .

$n = 100, \phi = 0.5$	$\mu$	$\delta_1$	$\phi$	$\sigma$	$\mu$	$\delta_1$	$\phi$	$\sigma$
	$p = 2.0$				$p = 2.5$			
Info. matrix	0.58	0.35	0.51	1.42	0.79	0.44	0.60	1.11
Simulated	0.66	0.31	0.52	3.30	0.87	0.42	0.62	2.05
	$p = 3.5$				$p = 5.0$			
Info. matrix	0.91	0.50	0.71	0.86	0.94	0.54	0.71	0.67
Simulated	1.04	0.50	0.71	0.93	1.03	0.50	0.75	0.67
	$p = 7.0$				$p = 10.0$			
Info. matrix	0.98	0.66	0.72	0.61	0.99	0.46	0.76	0.56
Simulated	1.04	0.66	0.79	0.65	1.17	0.48	0.79	0.59

The results for  $\delta_2$  are essentially the same as for  $\delta_1$ . For  $p = 2.0$  and  $2.5$ , the simulated values for  $\sigma$  are much different than those obtained from the sample information matrix but that is due to the fact that the population kurtosis is infinite; see also Islam and Tiku [3, p. 2453].

$$-\frac{\partial^2 \ln L}{\partial \sigma^2} = \frac{2p}{k\sigma^2} \left[ \sum_{i=1}^n 2z_i g(z_i) + \sum_{i=1}^n z_i^2 h(z_i) \right] - \frac{n}{\sigma^2}$$

where

$$g(z_i) = \frac{z_i}{1 + z_i^2/k}, \quad h(z_i) = \frac{1 - z_i^2/k}{(1 + z_i^2/k)^2}, \quad u_{ji} = x_{j,i} - \phi x_{j,i-1},$$

and

$$w_{i-1} = y_{i-1} - \delta_1 x_{1,i-1} - \delta_2 x_{2,i-1} - \cdots - \delta_q x_{q,i-1}.$$

The simulated variances and the corresponding values obtained from the sample information matrix are given in Table A.1. The agreement between the two is quite close, for large  $n$  ( $\geq 100$ ).

## References

- [1] J. Durbin, Estimation of parameters in time series regression model, *Journal of Royal Statistics Society B* 22 (1960) 139–153.
- [2] M.L. Tiku, W.K. Wong, G. Bian, Estimating parameters in autoregressive models in non-normal situations: Symmetric innovations, *Communications in Statistics Theory and Methods* 28 (1999) 315–341.
- [3] M.Q. Islam, M.L. Tiku, Multiple linear regression models under non-normality, *Communications in Statistics Theory and Methods* 33 (2004) 2443–2467.
- [4] D.C. Vaughan, M.L. Tiku, Estimation and hypothesis testing for a non-normal bivariate distribution and applications, *Journal of Mathematical and Computer Modelling* 32 (2000) 53–67.
- [5] M.L. Tiku, D.C. Vaughan, Logistic and nonlogistic density functions in binary regression with nonstochastic covariates, *Biometrical Journal* 39 (1997) 883–898.
- [6] D.C. Vaughan, The generalized secant hyperbolic distribution and its properties, *Communications in Statistics Theory and Methods* 31 (2002) 219–238.
- [7] A.D. Akkaya, M.L. Tiku, Robust estimation in multiple linear regression model with non-Gaussian noise, *Automatica* 44 (2008) 407–417.
- [8] M.L. Tiku, R.P. Suresh, A new method of estimation for location and scale parameters, *Journal of Statistical Planning and Inference* 30 (1992) 281–292.
- [9] M.L. Tiku, W.K. Wong, D.C. Vaughan, G. Bian, Time series models in non-normal situations: Symmetric innovations, *Journal of Time Series Analysis* 21 (2000) 571–596.
- [10] M.L. Tiku, M.Q. Islam, H.S. Sazak, Estimation in bivariate nonnormal distributions with stochastic variance functions, *Computational Statistical & Data Analysis* 52 (2008) 1728–1745.
- [11] S. Puthenpura, N.K. Sinha, Modified maximum likelihood method for the robust estimation of system parameters from very noisy data, *Automatica* 22 (1986) 231–235.
- [12] M.L. Tiku, Estimating the mean and standard deviation from a censored normal sample, *Biometrika* 54 (1967) 155–165.
- [13] M.L. Tiku, W.Y. Tan, N. Balakrishnan, Robust Inference, Dekker Inc., New York, 1986.
- [14] W.B. Smith, C.D. Zeis, G.W. Syler, Three parameter lognormal estimation from censored data, *Journal of Indian Statistical Association* 11 (1973) 15–31.
- [15] D.R. Cox, Further results on tests of separate families of hypotheses, *Journal of Royal Statistical Society Series B* 24 (1962) 406–423.
- [16] B. Şenoğlu, M.L. Tiku, Censored and truncated samples in experimental design under non-normality, *Statistical Methods* 6 (2004) 173–199.
- [17] R.A. Maller, Regression with autoregressive errors—some asymptotic results, *Statistics* 20 (1989) 23–39.
- [18] K.O. Cogger, Robust time series analysis—an  $L_1$  approach, in: K.K. Lawrence, J.L. Arthur (Eds.), *Robust Regression*, Marcel Dekker, New York, 1990.
- [19] S. Schäffler, Maximum likelihood estimation for linear regression model with autoregressive errors, *Statistics* 22 (1991) 191–198.
- [20] N.K. Nagaraja, W.A. Fuller, Least squares estimation of the linear model with autoregressive errors, in: *New Directions in Time Series Analysis, Part I*, in: IMA Math. Appl., vol. 45, Springer, New York, 1992, pp. 215–225.
- [21] H.D. Vinod, L.R. Shenton, Exact moments for autoregressive and random walk models for a zero or stationary initial value, *Econometric Theory* 12 (1996) 481–499.
- [22] A.D. Akkaya, M.L. Tiku, Estimating parameters in autoregressive models in non-normal situations: Asymmetric innovations, *Communications in Statistics Theory and Methods* 30 (3) (2001) 517–536.
- [23] A.D. Akkaya, M.L. Tiku, Corrigendum: Time series models with asymmetric innovation, *Communications in Statistics Theory and Methodology* 30 (10) (2001) 2227–2230.
- [24] M.L. Tiku, A.D. Akkaya, Robust Estimation and Hypothesis Testing, New Age International (P) Ltd., New Delhi, 2004, Also, Oscar Publications, [www.alltimebooks.com](http://www.alltimebooks.com).
- [25] W.Y. Tan, V. Lin, Some robust procedures for estimating parameters in an autoregressive model, *Sankhya B* 55 (1993) 415–435.
- [26] D.J. Hand, F. Daly, A.D. Lunn, K.J. McConway, E. Ostrowski, *Small Data Set*, Chapman and Hall, New York, 1994.
- [27] M.L. Tiku, M.Q. Islam, A.S. Selçuk, Non-normal regression II, *Communications in Statistics Theory and Methods* 30 (2001) 1021–1045.
- [28] A.D. Akkaya, M.L. Tiku, Robust estimation and hypothesis testing under short-tailedness and inliers, *Test* 14 (2005) 129–150.
- [29] E.S. Pearson, M.L. Tiku, Some notes on the relationship between the distributions of central and noncentral F, *Biometrika* 57 (1970) 175–179.
- [30] L.C. Hamilton, *Regression with Graphics*, Brooks/Cole Publishing Company, CA, 1992.